

# No-Corpus Artificial Intelligence from Algebraic Abiogenesis: Policy-Governed Object Formation, Theory Ecologies, and Certified Memory

Douglas C. Youvan

[doug@youvan.com](mailto:doug@youvan.com)

[www.youvan.ai](http://www.youvan.ai)

July 11, 2026

Corpus-trained artificial intelligence learns primarily by compressing human-produced examples into statistical parameters. Algebraic Abiogenesis suggests a different cognitive architecture. A no-corpus system begins with finite operational worlds, generates its own experiments, detects exact or controlled invariances, forms equivalence classes of interchangeable presentations, extracts compact theories, and challenges those theories with countermodels. Its persistent knowledge is not merely a parameter vector but an auditable ecology of objects, laws, proofs, failures, and refinement histories. Formally, an operational ecology  $E$  induces an observational congruence  $\Theta_E$  on a free expression algebra  $T$ . The quotient  $T/\Theta_E$  is the system's provisional ontology. A basis  $\Sigma$  compresses the observed equivalences, while the residual theory  $R(E|\Sigma)$  identifies unexplained structure. Countermodels refine  $\Theta_E$  and prevent premature conceptual collapse. This paper develops Algebraic Abiogenesis as a candidate cognitive core for artificial intelligence without inherited corpus training. It defines corpus independence, policy-governed objecthood, theory-forming learning, certified memory, intrinsic discovery objectives, and a feasible finite graph implementation. The proposal does not eliminate data, computation, or inductive bias. It replaces imitation of a fixed corpus with self-generated operational experience and explicit theory revision.

Keywords: no-corpus artificial intelligence, Algebraic Abiogenesis, Mathematical It from Bit, object formation, quotient learning, theory ecology, automated discovery, certified memory, countermodel refinement, intrinsic motivation, artificial scientist.

A collaboration with OpenAI GPT-5.5 Thinking. CC4.0.

<https://orcid.org/0009-0004-4206-4253>

## 1. Introduction

Modern artificial intelligence is dominated by corpus training. A model receives large collections of text, images, audio, code, demonstrations, or labeled examples. Optimization modifies a high-dimensional parameter system so that the model predicts missing, subsequent, or desired outputs.

This paradigm is powerful, but it places human-produced representations before machine intelligence. The model begins inside an inherited conceptual world. Its vocabulary, distinctions, examples, errors, conventions, and preferred objects are already embedded in the corpus.

Algebraic Abiogenesis begins from a different order of construction. Anonymous finite operations are supplied before named algebraic objects or target equations. Formal expressions are generated, evaluated, compared, compressed into laws, and challenged by countermodels. Stable equivalence classes become quotient objects. Compact bases become theories.

The technical foundation was developed in *Mathematical It from Bit: Algebraic Abiogenesis, Anonymous Finite Operations, and the Automated Birth of Abstract Algebra*. A public explanation was provided in *How Mathematics Can Be Born: A Public Guide to Algebraic Abiogenesis and Mathematical It from Bit*. The associated code, finite tables, machine-readable records, and reproduction tests are publicly archived.

The present paper extends that framework from automated algebraic discovery to artificial intelligence architecture.

The central proposal is:

A future artificial intelligence may acquire its core ontology without inherited corpus training by generating operational worlds, discovering admissible equivalences, forming quotient objects, extracting theories, and preserving certified refinement histories.

This is not intelligence without data. It is intelligence without a fixed external corpus serving as the primary source of concepts.

## 2. Corpus Independence

Definition 2.1. A system is corpus-trained when its central representational state is optimized using a large externally supplied collection of human-produced or externally curated examples.

Definition 2.2. A system is core corpus-independent when its principal objects, laws, and internal theories are generated from:

1. self-generated operational worlds;
2. direct environmental interaction;

3. internally generated experiments;
4. exact or explicitly bounded verification;
5. fixed meta-policies that do not encode the target theory.

A corpus-independent system may still receive sensor data. It may communicate with humans. It may later learn language conventions. Corpus independence concerns the origin of its core ontology and reasoning structures, not the absence of all input.

A language interface cannot learn English without exposure to English. A robot cannot learn its physical environment without observations. A mathematical discovery engine cannot evaluate an operation without a table or executable rule.

The distinction is therefore:

external corpus imitation versus internally generated theory formation.

The conventional learning pipeline is

corpus

→ optimization

→ statistical representation

→ prediction.

The proposed pipeline is

operational world

→ experiment

→ invariance

→ admissible equivalence

→ quotient object

→ theory

→ prediction and intervention.

The second pipeline does not require a named target object. It requires only a substrate, an experiment language, an evaluation procedure, and general discovery policies.

### **3. Algebraic Cognitive State**

Let  $\tau$  be a finite operational signature and let  $X$  be a finite variable set. Let

$T_{\tau}(X)$

be the absolutely free term algebra generated by  $X$ .

At time  $t$ , the system maintains an ecology

$$E_t = \{A_1, \dots, A_k\}$$

of finite operational worlds interpreting  $\tau$ .

Each  $A \in E_t$  induces an evaluation homomorphism

$$\phi_A \tau(X) \rightarrow A^A(A^X).$$

Define the observational congruence

$$\Theta_t$$

$$\bigcap_{A \in E_t} \ker \phi_A.$$

Thus

$$s \Theta_t u$$

when  $s$  and  $u$  behave identically under every assignment in every current operational world.

Definition 3.1. The provisional ontology at time  $t$  is

$$Q_t$$

$$\tau(X) / \Theta_t.$$

The elements of  $Q_t$  are equivalence classes of expressions. Two presentations count as one object when the current ecology cannot distinguish them under the declared tests.

The ontology is neither a vocabulary list nor an embedding space. It is an explicit quotient.

Definition 3.2. A policy  $\pi$  specifies which observations, transformations, interventions, tasks, or consequences are relevant to identity.

The policy-indexed congruence is written

$$\Theta_t, \pi.$$

Different policies may produce different object systems from the same operational substrate.

For example, two states may be equivalent for prediction but not for intervention. Two expressions may be equivalent under one family of transformations but separable under another.

The useful regime lies between two failures.

If

$$\Theta t, \pi = \Delta,$$

where  $\Delta$  is literal syntactic identity, then no nontrivial abstraction has formed.

If

$$\Theta t, \pi = T\tau(X)^2,$$

then all expressions have collapsed into one object.

A viable ontology requires

$$\Delta \subsetneq \Theta t, \pi \subsetneq T\tau(X)^2.$$

The system must forget some differences without forgetting every difference.

#### 4. Concept Formation as Quotient Formation

Definition 4.1. A machine concept is a policy-stable equivalence class of presentations.

This definition makes conceptual identity operational.

A concept is not initially a human label. It is a set of states, expressions, trajectories, or models that may substitute for one another without changing the consequences relevant under  $\pi$ .

Suppose several sensory configurations transform identically under the system's available actions. They may form one object class.

Suppose several programs produce the same observable behavior under all current tests. They may form one computational concept.

Suppose several algebraic expressions induce the same function in every model of an ecology. They become one quotient element.

Object formation is therefore controlled loss of distinction.

The process is

presentation

→ operational testing

→ stable interchangeability

→ equivalence class

→ object.

This gives a precise alternative to treating concepts as hidden vectors learned from statistical co-occurrence.

A vector representation may still be used for efficient search. It is not the final identity criterion. The final criterion is substitution under declared policies and tests.

Proposition 4.1. If  $E_t \subseteq E_{t+1}$  under a fixed policy and experiment language, then

$$\Theta_{t+1}, \pi \subseteq \Theta_t, \pi.$$

Proof. Adding operational worlds adds equality constraints to the intersection of evaluation kernels. An identity surviving the larger ecology must survive the smaller ecology. Therefore the later congruence is finer. ■

The result gives monotone observational refinement. New countermodels can split old concepts. They cannot merge previously separated expressions unless the policy, language, or observation set is changed.

Policy revision may intentionally coarsen or reorganize the ontology. Such changes must be recorded separately from evidence-driven refinement.

## 5. Theory-Forming Learning

A quotient records which expressions are identified. A theory explains why.

Let

$$K_d(E_t)$$

be the exact bounded theory observed through term complexity  $d$ .

Let  $\Sigma_t$  be a provisional basis of laws.

Let

$$Cnd(\Sigma_t)$$

be the bounded deductive closure of  $\Sigma_t$ .

Define the residual theory

$$Rd(E_t | \Sigma_t)$$

$$K_d(E_t) \setminus Cnd(\Sigma_t).$$

Learning proceeds on the residual, not on the total collision catalogue.

If a single shallow law explains thousands of observed equalities, those equalities cease to count as new discoveries. The system searches for remaining unexplained structure.

Definition 5.1. Theory-forming learning is the iterative update

$(E_t, \Theta_t, Q_t, \Sigma_t, R_t)$

$\rightarrow$

$(E_{t+1}, \Theta_{t+1}, Q_{t+1}, \Sigma_{t+1}, R_{t+1}),$

where each update is produced by new experiments, new worlds, new laws, new countermodels, or policy revision.

A minimal cycle is:

1. Generate operational worlds.
2. Generate expressions or interventions.
3. Evaluate them.
4. Detect exact or bounded equivalences.
5. Form a quotient ontology.
6. Extract a compact explanatory basis.
7. Subtract explained consequences.
8. Search the residual.
9. Generate countermodels.
10. Refine the ecology and ontology.
11. Store the result with provenance.

This differs from ordinary gradient training.

The primary persistent change is not a weight update. It is a revision of objects, laws, and their evidential status.

The system may still use neural, evolutionary, or numerical components to search large spaces. Those components serve the theory-forming process rather than constituting the complete memory of the system.

## 6. Countermodels and Endogenous Criticism

A no-corpus intelligence requires a mechanism stronger than confirmation.

Let  $\Sigma$  be a provisional basis and let  $p$  be an observed identity.

A countermodel is an operational world  $B$  satisfying

$B \models \Sigma$

but

$B \neq p$ .

The system deliberately searches for such B.

If B is found, it is added to the ecology. The observational congruence refines, the quotient may split, and p is removed from the general theory.

This produces endogenous scientific criticism.

The AI does not merely collect examples supporting its current concepts. It constructs worlds designed to destroy them.

Countermodels perform four cognitive functions.

First, they prevent finite coincidence from becoming theory.

Second, they distinguish a general theory from specialized branches.

Third, they reveal which laws are independent.

Fourth, they recover distinctions erased by overaggressive quotienting.

A corpus-trained model usually receives criticism through additional data, human feedback, or revised objectives. A theory-forming system generates a substantial part of its criticism internally.

Its learning cycle is therefore adversarial in an epistemic rather than merely competitive sense.

## **7. Intrinsic Objectives Without Target Leakage**

A no-corpus system still requires goals. Absence of a corpus does not imply absence of inductive bias.

The central restriction is that the objective must not encode the desired theory or object.

Rewarding the discovery of associativity would bias the system toward semigroups. Rewarding a Mandelbrot equation would predetermine the fractal. Rewarding a known scientific law would convert discovery into recovery.

Acceptable objectives are meta-theoretical.

Examples include:

1. increase exact predictive compression;
2. reduce unexplained residual theory;

3. preserve nondegenerate distinctions;
4. increase transfer across independently generated worlds;
5. maximize counterexample value;
6. prefer shorter explanations;
7. improve intervention success;
8. reduce computational cost;
9. preserve auditable provenance.

A generic experiment value may be written

**V(e)**

$\alpha\Delta U$

+

$\beta\Delta R$

+

$\gamma C$

+

$\delta T$

–

$\lambda\text{Cost}(e),$

where:

$\Delta U$  is expected uncertainty reduction;

$\Delta R$  is expected residual-theory reduction;

$C$  is expected countermodel value;

$T$  is expected transfer value.

The coefficients are meta-policy parameters. They do not identify the target law.

Collapse penalties are required. Otherwise the system may maximize compression by making all states equivalent.

The objective should therefore reward structured compression, not maximal compression.

A valid discovery objective encourages the system to find laws without specifying which laws it must find.

## 8. Certified Cumulative Memory

A conventional model stores most acquired capability diffusely in parameters. A theory-forming system requires explicit memory objects.

Definition 8.1. A certified memory item is a tuple

**M**

$(\tau, E, \pi, \Sigma, Q, P, C, S),$

where:

$\tau$  is the operational language;

$E$  is the supporting ecology;

$\pi$  is the identity policy;

$\Sigma$  is the theory basis;

$Q$  is the quotient ontology;

$P$  is the proof or verification record;

$C$  is the countermodel history;

$S$  is the epistemic status.

Possible status values include:

observed;

recurrent;

conjectural;

proved;

counterexample-bounded;

known rediscovery;

candidate-original;

rejected;

superseded.

The memory system must preserve negative knowledge.

Failed laws, collapsing ontologies, unsuccessful searches, and separating countermodels are not discarded. They constrain future search.

The resulting memory is a theory ecology.

Nodes are objects, theories, models, and countermodels.

Edges record:

derivation;

refinement;

specialization;

reduct;

expansion;

interpretation;

contradiction;

and historical equivalence.

Learning becomes growth and revision of this ecology.

The memory can be audited because every promoted claim points to:

the operational worlds supporting it;

the terms or experiments involved;

the laws explaining it;

the countermodels attempted;

and the verification procedure used.

This does not guarantee correctness. It makes the route to the claim inspectable.

## **9. Minimal No-Corpus Architecture**

A minimal implementation contains nine modules.

### **9.1 Substrate generator**

Produces finite operational worlds.

Suitable substrates include:

finite algebras;  
directed graphs;  
state-transition systems;  
Boolean networks;  
rewrite systems;  
small programs;  
and bounded simulated environments.

### **9.2 Experiment generator**

Builds terms, transformations, paths, compositions, interventions, and tests.

### **9.3 Exact evaluator**

Computes the outcome of every experiment in every finite world whenever exhaustive evaluation is feasible.

### **9.4 Invariant detector**

Finds expressions or states that remain indistinguishable under the current tests.

### **9.5 Quotient constructor**

Forms provisional objects as equivalence classes.

### **9.6 Theory extractor**

Compresses observed equivalences into compact laws.

### **9.7 Residual search engine**

Removes explained consequences and searches for additional structure.

### **9.8 Countermodel engine**

Constructs operational worlds intended to falsify current conjectures.

### **9.9 Certified memory manager**

Stores theories, quotients, proofs, failures, and lineage records.

A numerical search module may propose worlds or experiments. A language model may help translate discoveries into human prose. Neither must define the core ontology.

The architecture is closer to an artificial scientist than to a conversational predictor.

## 10. A Feasible Finite Graph Chamber

A serious first implementation can be built on one commodity computer.

The substrate consists of finite directed graphs with anonymous local transformations.

No text corpus is used.

No pretrained model is required.

No target graph family is supplied.

A world contains:

a finite directed graph;

a finite set of anonymous rewrite or transition operations;

a bounded intervention language;

and an exact state evaluator.

The system performs the following loop.

1. Generate graphs of several orders.
2. Generate anonymous graph transformations.
3. Compose transformations into bounded experiments.
4. Evaluate all reachable outcomes.
5. Detect behaviorally equivalent states or transformation expressions.
6. Form quotient state classes.
7. Extract short laws governing those classes.
8. Subtract the consequences of those laws.
9. Cluster worlds by residual theory.
10. Generate adversarial graphs and transformations.
11. test transfer at larger graph orders.
12. Issue birth certificates for surviving objects and theories.

A discovered object is promoted only when it:

recurs across independently generated graph worlds;

survives order transfer;  
improves prediction or intervention;  
admits compact explanation;  
and survives adversarial refinement.

The output is not primarily an image or verbal response. It is a certified theory package.

A record may contain:

Substrate:

finite directed graph dynamics

Policy:

equivalence under declared interventions

Object:

quotient class  $Q$

Basis:

$\Sigma$

Observed theory:

$K$

Residual:

$K \setminus Cn(\Sigma)$

Countermodels:

listed finite graph worlds

Transfer:

verified at graph orders  $n_1, \dots, n_k$

Status:

calibration, candidate-original, or rejected

This would constitute a real no-corpus concept-formation experiment.

It would not constitute general intelligence.

### **11. Evaluation Criteria**

The architecture must be compared with simpler systems.

Relevant baselines include:

uniform random world generation;

collision-only search;

fixed-world invariant detection;

direct law enumeration;

neural predictive compression without quotient certification;

and ablations removing countermodels, residual subtraction, or policy control.

A successful system should demonstrate:

1. formation of nontrivial quotient objects;
2. recurrence across world sizes;
3. compact explanatory laws;
4. reduction of residual uncertainty;
5. adversarial survival;
6. transfer to withheld worlds;
7. predictive or intervention advantage;
8. reproducibility;

9. explicit failure retention;

10. no target theory encoded in the objective.

The strongest result would show that discovered objects improve performance in worlds not used during their formation.

Recognition by humans is secondary.

A concept earns its status by operational usefulness and theoretical stability.

## **12. Limitations**

The proposal does not remove prior structure.

The designer still supplies:

a substrate family;

an experiment language;

an equality or comparison procedure;

computational limits;

and discovery policies.

The method therefore does not create intelligence from literal nothing.

Finite exact evaluation also scales poorly. Larger worlds require sampling, abstraction, symbolic execution, or approximate search.

Approximate equivalence creates an additional problem. Exact congruences compose cleanly. Statistical similarities may fail under context. A future system must distinguish heuristic clustering from identity licensed for substitution.

Policy selection is another unresolved issue. A poor policy may produce useless objects even when the finite computations are exact.

Historical novelty remains difficult. A machine may independently generate a theory already known to humans.

Conceptual significance is not reducible to compression alone.

Language remains external to the core architecture. Human communication requires exposure to linguistic conventions, even if the underlying ontology was generated independently.

Finally, the current Algebraic Abiogenesis demonstrations are calibrations in finite algebra. Their extension to sensorimotor intelligence, causal discovery, or open physical environments remains a research hypothesis.

### **13. Conclusion**

Corpus-trained intelligence begins with the accumulated representational surface of human culture.

A no-corpus Algebraic-Abiogenesis system would begin below that surface.

It would generate operational worlds, perform experiments, detect stable equivalences, form quotient objects, extract theories, manufacture countermodels, and preserve certified memory.

Its core learning transition would be

difference

→ recurrence

→ equivalence

→ object

→ law

→ theory

→ action.

The proposal does not reject neural computation, optimization, or language models. It assigns them subordinate roles within a larger theory-forming architecture.

The central persistent state is not only a parameter vector.

It is a revisable ecology of objects and theories.

The deepest learning operation is not weight adjustment.

It is controlled quotient revision.

Algebraic Abiogenesis therefore provides a candidate cognitive core for future artificial intelligence without inherited corpus training: policy-governed object formation, explanation-subtracted theory discovery, adversarial refinement, and certified cumulative memory.

The decisive next experiment is not conversational fluency.

It is the autonomous formation of a useful object that was not encoded in the substrate, objective, examples, or vocabulary of the system.

## References

1. Youvan, D. C. *Mathematical It from Bit: Algebraic Abiogenesis, Anonymous Finite Operations, and the Automated Birth of Abstract Algebra*. July 2026. DOI: 10.13140/RG.2.2.25987.26403.
2. Youvan, D. C. *How Mathematics Can Be Born: A Public Guide to Algebraic Abiogenesis and Mathematical It from Bit*. July 2026. DOI: 10.13140/RG.2.2.30600.99847.
3. Youvan, D. C. *Algebraic Abiogenesis: Methodology, Code, Finite Operation Tables, and Reproducibility Records*. GitHub repository, 2026.  
<https://github.com/DougYouvan/algebraic-abiogenesis>.

The author has published over 5,700 AI-assisted papers at:

<https://www.researchgate.net/profile/Douglas-Youvan/research> . Google Scholar has indexed these preprints, and if you want a particular reference, the AI mode of a Google search (Gemini) has efficiently catalogued these preprints.